Section 17

Variations on Random Sampling

As noted in Section 3, a *population* consists of all members of the group of interest to a researcher. A population may be small, such as all social workers employed by a public hospital in Detroit, or it may be large, such as all social workers in Michigan. The larger a population, the more likely a researcher will study only a sample of the population and *infer* that what is true of the sample is also true of the population.[1] The process of making such an inference is referred to by statisticians as *generalizing* from a sample to a population.

Freedom from bias is the most important characteristic of a sample.[2] An unbiased sample is defined as one in which all individuals in a population have an equal chance of being included as a participant. The basic method for obtaining an unbiased sample is to use *random sampling* from a population. Various types of random sampling are described below.

Random sampling yields an unbiased sample.

To draw a *simple random sample*, a researcher can put names on slips of paper and draw the number needed for the sample. This method is efficient for drawing samples from small populations.

Names can be drawn to obtain a random sample.

To draw a simple random sample from a large population, it is more efficient to use a *table of random numbers* than to write names on slips of paper. A portion of a table of random numbers is shown in Appendix C near the end of this book.[3] In this table, there is no sequence to the numbers, and in a large table of random numbers, each number appears about the same amount of times. To use the table, first assign everyone in the population a *number name*. For instance, if there are 90 individuals in a population, name the first individual 01, the second individual 02, the third individual 03, and so on, until you reach the last individual, whose number is 90.[4] (Often, computerized records have the individuals already numbered, which simplifies the process. Any set of numbers will work as number names as long as each individual has a different number and all individuals have the same number of digits in their number names.) To use the table, flip to any page in a book of random numbers and put your finger on the page without looking. This will determine the starting point. Let us

A *table of random numbers* may also be used to draw a random sample.

To use a *table of random numbers*, give each person in the population a *number name*.

start in the upper left-hand corner of the table in Appendix C for the sake of illustration. Because each individual has a two-digit number name, the first two digits identify the first participant; this is individual number 21. The next two digits to the right are 0 and 4; thus, individual number 04 will also be included in the sample. The third number is 98. Because there are only 90 in the population, skip 98 and continue to the right to 08, which is the number of the next individual drawn. Continue moving across the rows to select the sample.

Stratified random sampling is usually superior to simple random sampling. In stratified random sampling, the population is first divided into strata that are believed to be relevant to the variable(s) being studied. Suppose, for instance, you wanted to survey opinions on alcohol consumption with all students on a college campus as the population. If you suspect that men and women might differ in their opinions on alcohol consumption, it would be desirable to first stratify the population according to gender and then draw separately from each stratum at random. Specifically, you would draw a random sample of men and then separately draw a random sample of women. The same percentage should be drawn from each stratum. For instance, if you want to sample 10% of the population and there are 1,600 men and 2,000 women, you would draw 160 men and 200 women. Notice that there are more women in the sample than men, which is appropriate because the women are more numerous in the population. It is important to note that you would *not* be stratifying in order to compare men with women. Rather, the purpose of stratifying is to obtain a single sample of the college population that is representative in terms of gender.[5]

In *stratified random sampling*, draw participants at random separately from each stratum.

Draw the same percentage, not the same number, from each stratum.

In the stratified random sample being considered, the benefits of randomization (i.e., the elimination of bias) have been retained. In addition, you have gained the advantage of having appropriate proportions of men and women in the sample. If your hunch was correct that men and women differ in their opinions on alcohol consumption, you would have increased the precision of the results by stratifying.[6]

Note that stratifying does not eliminate all sampling errors. For instance, when you drew the women at random, you may have by chance obtained women for the sample who are not representative of all women on the campus; the same, of course, holds true for men. Nevertheless, stratifying has eliminated the possibility of obtaining a disproportionately large number of either men or women for the sam-

(i.e., it eliminates a particular *type* of sampling error—*not* all sampling errors).

For large-scale studies, *multistage random sampling* may be used. In this technique, a researcher might do the following. Stage 1: Draw a sample of counties at random from all counties in a state; Stage 2: Draw a sample of voting precincts at random from all precincts in the counties previously selected; and Stage 3: Draw individual voters at random from all precincts that were sampled. In multistage sampling, a researcher could introduce stratification. For instance, a researcher could first stratify the counties into rural, suburban, and urban and then separately draw counties at random from these three types of counties, thereby ensuring that all three types of counties are included in the sample.

Multistage random sampling may be used in large-scale studies.

A technique that is sometimes useful is *random cluster sampling*. For use of random cluster sampling, all members of a population must belong to a cluster (i.e., an existing group). For example, all Boy Scouts belong to troops; in most high schools, all students belong to homerooms; and so on. Unlike simple random sampling, in which individuals are drawn, in cluster sampling, *clusters* are drawn. To conduct a survey of Boy Scouts, for instance, a researcher could draw a random sample of troops, contact the leaders of the selected troops, and ask them to administer the questionnaires to all members of their troops.

In *random cluster sampling*, existing groups of participants are drawn.

There are two advantages to random cluster sampling over the other two types of random sampling already discussed. First, there are fewer individuals for a researcher to contact (e.g., only the troops' leaders and not the individual scouts). Second, the degree of cooperation is likely to be greater if troop leaders ask the individual scouts to participate than if a researcher who is unknown to the scouts asks them.

There is also a disadvantage to random cluster sampling. For statistical reasons that are beyond the scope of this book, the number of clusters (not the number of participants) should be treated as the sample size. Thus, for instance, if 20 troops with 10 Boy Scouts each were selected, responses from 200 Boy Scouts would be obtained. However, the researcher would need to report the sample size as 20, not as 200. As a result, when using cluster sampling, it is desirable to use a large number of clusters to overcome the disadvantage.

As you know from Section 3, random sampling creates *sam-*

pling errors. The inferential statistics described in this part of the book are designed to assess the impact of these errors on the results researchers obtain when they use random samples.

Endnotes

[1] This part of the book (Part D) describes *inferential statistics*, which are statistics that help assess the appropriateness of making *inferences* from a sample to a population.
[2] Examples of biased samples are provided in Section 3 of this book.
[3] Academic libraries have books of random numbers. Statistical computer programs can also generate them.
[4] The number of digits in the number names must equal the number of digits in the population total. For instance, if there are 500 people in a population, there are 3 digits in the total and there must be 3 digits in each number name. Thus, the first case in the population is named 001, the second one is named 002, and so on.
[5] If your purpose was to compare men with women, then it would be acceptable to draw the same number of each and compare averages or percentages for the two samples.
[6] Of course, if your hunch that men and women differ in their opinions was wrong, the use of stratification would be of no benefit.

Exercise for Section 17

Note: If any of your answers include the term *random sampling*, indicate whether it is *simple*, *stratified*, *cluster*, or *multistage*.

Factual Questions

1. The most important characteristic of a good sample is that it is free from what?
2. If you put the names of all members of a population on slips of paper, mix them, and draw some, what type of sampling are you using?
3. If there are 60 members of a population and you give them all number names starting with 01, what are the number names of the *first two participants selected* if you select a sample starting at the beginning of the third row of the Table of Random Numbers in Appendix C near the end of this book?
4. If there are 500 members of a population and you give them all number names starting with 001, what are the number names of the *first two participants selected* if you select a sample starting at the beginning of the fourth row of the Table of Random Numbers in Appendix C near the end of this book?